Research Article Vol. 11, No. 6 / June 2024 / Optica 759

Multi-contrast x-ray identification of inhomogeneous materials and their discrimination through deep learning approaches

T. Partridge,[1,*] S. S. Shankar,[2] I. Buchanan,[1] P. Modregger,[3] A. Astolfo,[1,4] D. Bate,[1,4] and A. Olivo[1]

[1]Department of Medical Physics and Biomedical Engineering, University College London, London WC1E 6BT, UK
[2]Nylers Ltd., Marshall House, Middleton Road, Morden, Surrey SM4 6RW, UK
[3]Department of Physics, University of Siegen, 57072 Siegen, Germany
[4]Nikon X-Tek Systems Ltd., Tring, Herts HP23 4JX, UK
*tom.partridge@ucl.ac.uk

Received 2 October 2023; revised 27 March 2024; accepted 5 April 2024; published 23 May 2024

Recent innovations in x-ray technology (namely phase-based and energy-resolved imaging) offer unprecedented opportunities for material discrimination; however, they are often used in isolation or in limited combinations. Here we show that the optimized combination of contrast channels (attenuation at three x-ray energies, ultra-small angle scattering at two, standard deviation of refraction) significantly enhances material identification abilities compared to dual-energy x-ray imaging alone, and that a combination of off-the-shelf machine learning approaches can effectively discriminate, e.g., threat materials, in complex datasets. The methodology is validated on a range of materials and image datasets that are both an order of magnitude larger than those used in previous studies. Our results can provide an effective methodology to discriminate, and in some cases identify, different materials in complex imaging scenarios, with prospective applications across the life and physical sciences. While the detection of threat materials is used as a demonstrator here, the methodology could be equally applied to, e.g., the distinction between diseased and healthy tissues or degraded vs. pristine materials.

https://doi.org/10.1364/OPTICA.507049

1. INTRODUCTION

The introduction of phase-based methods has transformed x-ray imaging. These methods create contrast based on the unit decrement of the refractive index (δ if this is written as $n = 1 - \delta + i\beta$, with i being the imaginary unit) instead of the imaginary part, β, which determines the widely used attenuation coefficient by $\mu = 4\pi\beta/\lambda$ (with λ being the x-ray wavelength). Following pioneering experiments in the 1960s [1] and 1980s [2], the field exploded in the 1990s with the advent of 3rd generation synchrotron facilities [3,4] and with the first pioneering experiments based on more conventional micro-focal sources [5,6]. Effort was soon focused on approaches to perform quantitative phase imaging by separating attenuation and phase contributions (phase retrieval), with methods based both on crystals [7] and free-space propagation [8] being developed. Crystal-based methods, as well as methods developed later based on x-ray masks [9], gratings [10], or other beam modulators (e.g., [11–13]), are actually sensitive to x-ray refraction, i.e., to the first derivative of the phase shifts, which is linked to the refraction angle α by $\alpha = \frac{\lambda}{2\pi}\nabla_{x,y}\phi(x, y, \lambda)$, with $\nabla_{x,y}$ being the two-directional gradient operator transverse to the x-ray propagation direction z, and Φ being the phase shift introduced by the object. For this reason, these approaches are referred to as "differential phase" methods; if needed, the phase shift Φ can be obtained by integration.

Experimentation with crystal-based methods soon revealed that a third contrast channel could be retrieved, related to multiple refraction events caused by structures smaller than the spatial resolution of the imaging system. Early papers called this contrast channel refractive scattering [14] or extinction [15], or borrowed the term ultra-small angle x-ray scattering (USAXS) from the x-ray scattering community [16], although it must be borne in mind that in this case angles are in the order of microradians rather than degrees or tenths of a degree. Later, the community converged on the name dark-field [17], although USAXS is still widely used. The dark-field channel turned out to be accessible through practically all differential phase methods [17–19], including those adapted for use with low-brilliance x-ray sources in the 2000s [20,21], with options to access it also through propagation-based approaches emerging only more recently [22]. In addition to enabling some of the earliest clinical trials with phase-based x-ray methods [23], access to the dark-field alongside differential phase

with low-brilliance sources makes "multi-contrast" (phase, attenuation, dark-field) x-ray imaging readily available in standard labs, opening the way to commercial translation.

Edge illumination (EI), the phase-based method used in this work, uses apertured masks to generate phase sensitivity (see Section 2), an additional consequence of which is that the spatial resolution is determined by the size of the apertures in the pre-sample mask [24]. The dark-field signal does not depend directly on the relation between the size of the scatterers and size of the mask apertures [25]; however, in 2D (projection) images like the ones presented in this paper, it does depend on it indirectly. This dependence arises from how many edges the photons encounter on their path, and as the scatterers become larger, there will be fewer of them in a given sample thickness. Conversely, another contrast channel exploited in this work, the standard deviation of refraction, kicks in when the scatterers become larger than the aperture because sub-resolution features typically do not create a distinct refraction signal [25]. In this sense, it performs a similar role as the dark-field signal for features above the spatial resolution of the system (i.e., the pre-sample mask aperture).

Another transformative development in x-ray imaging was the introduction of energy-resolved detectors [26]. While solutions based on photon counters with thresholding capabilities are becoming increasingly popular [27,28], solutions exist that can provide full spectroscopic capabilities, such as the HEXITEC detector [29], pnCharge-Coupled Devices [30,31], and Controlled-Drift Detectors [32], albeit possibly at the cost of a reduced incident flux.

In this work, we used a photon counter with thresholding capabilities (see Section 2) that allows us to split the incoming x-ray spectrum in two parts; its use in phase-based multi-contrast imaging therefore makes the above three channels (plus potentially the standard deviation of refraction) simultaneously available in high and low energy versions, making six (or even eight) contrast channels available in principle. Furthermore, with mask-based methods like EI, the x-rays transmitted through the highly absorbing mask septa can also be collected separately to create a third attenuation image at a much higher x-ray energy [33]. Clearly, these contrasts are not all independent of each other. The independence of low and high energy x-ray attenuation images can be assumed when these are dominated by the photoelectric and the Compton effects, respectively [34]. More recently, a different energy dependence has also been observed in gratings-based dark-field imaging for features with sizes above and below a certain characteristic length of the imaging system [35]; however, this does not apply to EI [25]. Phase scales in a predictable way with x-ray energy, and its correlation with the Compton signal has been widely discussed (e.g., [36]). However, in the framework of attenuation-based imaging, it has been repeatedly observed that the problem over-determination allowed by the availability of attenuation images at more than two energies leads to an increased precision in the results (e.g., [37]), and similar principles can be extended to phase-based imaging [33,38], which this paper expands on to explore options for material identification based on the simultaneous availability of multiple image contrasts.

As a demonstrator, we apply the method to the area of security inspections, namely to the identification of threat materials. The field of x-ray based security scans is well developed, with exhaustive summaries provided in various review papers (e.g., [39]). While long-established dual-energy scanners are still widely used [40], technologies such as computed tomography (CT) are being deployed at airports worldwide, while others such as x-ray diffraction (e.g., [41]) are experiencing early market entry. However, the dark-field and standard deviation of refraction signals probe a different property of matter, namely microstructure instead of atomic number/electron density and molecular structure (probed by CT and diffraction, respectively), therefore providing information that is complementary to existing methods. Furthermore, both these signals are provided simultaneously with attenuation, and therefore on top of the currently used dual-energy images if an energy-resolved detector or appropriate filtration is used.

The combination of dark-field with deep learning approaches is also explored in this work. The popularity of deep learning methods has been booming in recent years [42], and their use in security applications is also rapidly expanding [43]. The prospect of automated detection is particularly attractive in security inspections, and indeed 3D imagery like that provided by CT is particularly well suited for analysis via machine learning methods [44]. Here we are interested in exploring the compatibility between multi-contrast x-ray imaging and deep learning, in particular the added value brought by the dark-field channel.

In the first part of the paper, we test the discrimination and identification potential of multi-contrast x-ray imaging used as a standalone analytic method with no added automatic detection algorithms. We apply this to a large dataset encompassing 19 threat and 56 non-threat materials of varying thickness, showing how the inclusion of additional contrast channels above conventional attenuation significantly aids material discrimination and often allows material identification. We propose a method to convert multi-contrast images into simpler, "material-specific" images and analyze the overall results obtained using this approach by means of truth tables. We extend the methodology to the case where different materials overlap, showing the approach still works.

In the second part of the paper, we create complex scenarios where the 75 materials are hidden inside bags and covered by a range of cluttering materials. Overall, we produced 3891 scans of randomly selected materials, with the thickness of each material varying between 12 and 30 mm in the various scans, combined with various cluttering objects. A total of 2732 (70%) of these scans were used as a training set (for transfer learning), and the remaining 1168 (30%) for testing. We progressively optimized our deep learning approach and obtained the best results with a hierarchical approach where the cluttering objects are segregated first then materials are discriminated; this was trained with a cross-entropy loss function and used layer-specific learning rate adaptation [45]. Despite the complexity of the dataset, use of this architecture resulted in a single false negative out of the 313 explosive-bearing cases contained in the set of 1168 images used for testing.

We note that, although here we used the discrimination of threat materials as a demonstrator, our interest lays in the general ability of the dark-field to separate and ideally identify materials using additional information regarding their microstructure.

2. METHODS

A. X-Ray Imaging System and Method

This work used the EI x-ray phase-contrast imaging method [9] in its laboratory implementation [21] to produce the dark-field and refraction images on top of the conventional attenuation ones. Combination with an x-ray detector with two energy thresholds

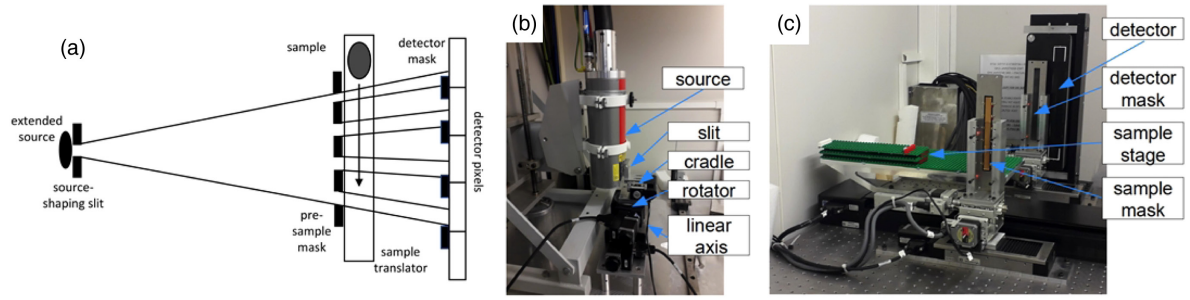

Fig. 1. (a) Schematic of the experimental setup, with (b) photos of the source and (c) detector/masks parts. A mismatch of the pre-sample mask position along the x-ray propagation axis allows for different detector apertures being hit at different points by the beamlets created by the pre-sample mask.

(XCounter XC-FLITE FX2, Direct Conversion, Danderyd, Sweden) allowed the acquisition of all the above images at two different average spectral energies. The x-ray source (MXR-160HP/11, Comet, Wünnewil-Flamatt, Switzerland) was set at 80 kVp and operated at 8.7 mA, and a higher detector threshold splitting the (W) spectrum with roughly one third and two thirds of the counts in the high and low energy bins, respectively. X-ray energies below approximately 24 keV were cut off by the lower detector threshold, so as to completely eliminate noise (i.e., detector pixels registered no counts with the beam off).

Two masks were used to realize the EI configuration. The first ("pre-sample") mask, placed immediately before the sample, had overall dimensions of 150 (h) × 9.6 (w) mm^2, with regularly spaced (75 μm period), 21.4 μm apertures parallel to the long side of the mask. The second ("detector") mask, placed immediately before the detector, had dimensions of 200 (h) × 12.8 (w) mm^2, 98 μm period, 28 μm apertures. The source-to-detector distance was 2.1 m; the detector mask was placed at 42 mm from the detector so that its projected period would match the 100 μm detector pixel pitch. The pre-sample mask was slightly displaced from its "ideal" position at 52.5 cm from the detector, where it also would also have matched the detector pitch; in this way, the beamlets it creates hit slightly different positions on the detector mask (see Fig. 1), allowing for the retrieval of attenuation, refraction, and dark-field with a single sample scan, similarly to the method described in [23]. Basically, different detector pixels receive different amounts of illumination [Fig. 1(a)], which correspond to sampling the illumination curve (IC) at different points, thus allowing its fit with an appropriate mathematical function. The IC is obtained by scanning the pre-sample mask while the detector mask is kept stationary, with apertures at the center of the corresponding pixels. This provides a bell-shaped curve with the maximum where pre-sample and detector mask apertures are aligned, and minima where the pre-sample mask apertures are aligned with the septa at either side of the corresponding detector mask aperture. The introduction of a sample causes the IC to dampen, shift laterally, and broaden, corresponding to attenuation, refraction, and dark-field effects, respectively. Since the IC is normally well approximated by a Gaussian function [18,33,46], fitting it before and after the introduction of a sample enables the extraction of the latter's attenuation, refraction, and dark-field characteristics on a pixel-by-pixel basis. Specifically, if

$$\mathrm{IC}_0 = \frac{a_0}{\sqrt{2\pi a_2}} \exp\left[\frac{(x-a_1)^2}{2a_2}\right] + a_3 \qquad (1)$$

is the IC without the object, in which a term (a_3) accounting for a degree of x-ray transmission through the mask septa has also been included, and

$$\mathrm{IC}_{\mathrm{obj}} = \frac{a_0 t}{\sqrt{2\pi(a_2+\sigma)}} \exp\left[\frac{(x-a_1-\Delta)^2}{2(a_2+\sigma)}\right] + a_3 o \qquad (2)$$

is the IC with the object, then t, Δ/d, σ/d^2, and o represent the transmission, refraction, dark-field, and offset images, respectively, with d the sample to detector mask distance used to transform the lateral shift of the IC into the refraction angle ϑ, in the approximation $\tan(\vartheta) \simeq \vartheta$.

Both masks are mounted on a pair of translators and a goniometer for alignment. An addition, a larger translator is used to scan the samples through the beam. An exposure time of 1 s was used, and samples were scanned at a speed of 0.3 mms^{-1}, meaning each scan took between 15 and 25 min, depending on the length of the specific sample. The relatively large (0.4 mm) focal spot of the source is reduced to approximately 80 μm in the horizontal direction by means of a Huber (Huber Diffraktionstechnik GmbH & Co. KG, Rimsting, Germany) slit, placed as close as possible to the source output window [see Fig. 1(b)]. A detailed description of the system can be found in [46], save for the use of a slightly different pre-sample mask. The entire system is placed inside a cabinet, which enables its easy transportation to different locations (e.g., for the explosive scans, see below).

B. Samples and Sample Preparation

A complete list of the materials considered in this application is provided in Supplement 1, Table 1, which also lists the cluttering objects used in the deep learning part of the study alongside the 57 benign and 19 explosive materials. The benign materials were scanned on the Nikon X-Tek Systems premises in Tring (Hertfordshire, UK) using the pre-commercial prototype described above. The prototype was then transported to the Cranfield Ordnance Test and Evaluation Centre (COTEC) near Devizes (Wiltshire, UK), which provided access to the explosive materials. These were first scanned individually as done for the benign materials in Tring and then were placed into bags and obscured by cluttering objects. A total of 3891 different combinations were produced and scanned, a description of which is provided in Supplement 1, Fig. 4, which also shows what combination of cluttering objects was used in combination with which material. At least 40 (typically more) cluttered combinations were realized for each of the threat materials, which means the overall

number of scans containing an explosive was approximately 22% of the total.

C. Deep Learning Architectures

The CNN Inception V3 [47] was used as our base architecture. This was pre-trained on ImageNet [48], with a few additional layers trained on our (training) dataset. For training, we used stochastic gradient descent with momentum as the optimization procedure, with L2 as the regularizer [48]. We trained for 20 epochs with a batch size of 64, using a dropout probability of 0.5 for the convolutional layers. A 10 times lower learning rate was used for the pre-trained layers in comparison to the additional ones, using batch normalization [49] after the convolutional layers of Inception V3.

Initially we used the softmax loss function, then introduced a cross-entropy loss. Assume we have N training samples, with M classes. Let the training set be represented by $\{(x_1, y_1), \ldots, (x_N, y_N)\}$ where x_i, $i \in \{1, \ldots, N\}$ represent N training images, and y_i denotes the corresponding ground-truth labels (since we have M classes, $y_i \in \{1, \ldots, M\}$). Training with a softmax loss is accomplished by minimizing the negative log-likelihood:

$$\mathcal{L}_s = -\frac{1}{N}\sum_{r=1}^{N} \log\left(\hat{p}_{r,y_r}\right),$$

where the probability $\hat{p}_{r,y_r}$, $r \in \{1, \ldots, N\}$ is obtained by applying the softmax function to the penultimate layer of the classifier. Letting $l_{r,m}$ denote the mth output for x_r, we have

$$p'_{r,m} = \frac{e^{l_{r,m}}}{\sum_{m'} e^{l_{r,m'}}}, \quad m, m' \in \{1, \ldots., M\}.$$

Conversely, with a sigmoid cross-entropy loss, the network is trained by minimizing the following loss function:

$$\mathcal{L}_e = -\frac{1}{NM}\sum_{r=1}^{N}[\boldsymbol{p_r} \cdot \log(\hat{\boldsymbol{p}}_r)] + [\,(1 - \boldsymbol{p_r}) \cdot \log(1 - \hat{\boldsymbol{p}}_r)],$$

with $\boldsymbol{p_r}$ being the M-dimensional ground truth probability vector, and $\hat{\boldsymbol{p}}_r$ being the M-dimensional predicted probability vector obtained by applying the sigmoid function to the outputs of the penultimate layer of the classifier.

While progressively refining the approach, the first improvement consisted of the introduction of a hierarchical architecture to segregate cluttering object classes first, then apply material discrimination. Assuming we have K object classes and R materials, an inception V3 net is first trained to detect one of the K objects $O_1 \ldots O_K$. K more inception V3 nets are then separately trained, where the lth net ($1 \le l \le K$) separates the benign and threat materials obscured by O_l. If a material image has been obscured by two objects O_{k_1} and O_{k_2}, the image is put into both O_{k_1} and O_{k_2} object classes during training with mild data augmentation (see Supplement 1, Fig. 4). During testing, the trained model chooses the k_1th material-discrimination net if the probability of the presence of object O_{k_1} is predicted to be greater than that of the presence of object O_{k_2}, and vice versa. This functionality naturally extends if a material is obscured by three or more objects.

The second improvement consisted of the introduction of the sigmoid cross-entropy loss function outlined above; note that the change of loss function only applied to the K nets trained for material discrimination, while the first net is still trained with the softmax loss. Finally, we introduced layer-specific learning rate adaptation by assigning a weight to a given layer proportional to the average decorrelation-based segregation between classes [50]; again, this was only applied to the K material-discrimination nets.

Training and inference procedures were carried out on an NVIDIA GeForce GTX 1080 Ti graphic card, with on-chip memory of ∼11 GB, which led to an inference time per image of the order of 10 ms.

3. RESULTS AND DISCUSSION

An initial simplified example of the benefit of using additional contrast channels alongside conventional ones is shown in Fig. 2. The horizontal axis in this plot is the new effective atomic number (Z_e), which is based on the effective atomic number (Z_{eff}) corrected using published linear x-ray attenuations [51] and represents a condensed version of dual energy attenuation-based imaging. The additional contrast channels accessed through phase contrast measurement with masks are then added on the vertical axis. These are dark-field at low [Fig. 2(a)] and high [Fig. 2(b)] energy, the "offset," i.e., the image at a much higher average x-ray energy obtained by exploiting x-rays that have traversed the highly absorbing mask septa [Fig. 2(c)], and the standard deviation of the refraction signal [Fig. 2(d)] which, as discussed above, can be used as a proxy for dark field when overlapping features larger than the system's spatial resolution are present [25,52]. For a straightforward comparison with Z_e, all contrast channels bar the standard deviation of refraction have been made thickness-independent through division by the attenuation signal, following the procedure described in [33]. Explosive materials are highlighted with a square for ease of visualization.

As can be seen, adding contrast channels separates out the points in the 2D graphs, allowing for much easier identification—especially of materials that would overlap based on their Z_e value alone. Unsurprisingly, the offset signal is the least effective in providing additional separation, due to its lack of complementarity over the two energies used to extract Z_e. Dark-field at low and high energy provide similar results, with the former spreading out the points a bit more due to the dark-field signal being stronger at lower energies [25,35]. The best separation seems to be provided by the standard deviation of refraction, which would indicate a prevalence of grain sizes comparable to or larger than the spatial resolution of the imaging system (approximately 25 μm).

From an operational perspective, the above information needs to be combined in a simplified way that could be presented to, e.g., an operator, without trading off on signal complexity. We propose a simple quantitative approach based on the relative distance between material pairs in all contrast channels, exemplified in Fig. 3. A radius corresponding to the measurement error [e.g., one standard deviation over a region of interest (ROI)] is assigned to all reference points (e.g., blue arrows in Fig. 3 for PE8). When an unknown material is scanned, contrast measurements and corresponding standard deviations are extracted from the corresponding ROI and plotted on top of the reference materials (red arrows in Fig. 3). While for obvious reasons only two contrast channels are represented in Fig. 3, this procedure actually defines a "hyper-cube" (or hyper-sphere if standard deviations are summed in quadrature), with dimensions corresponding to the number of used contrast channels. This immediately provides the probability

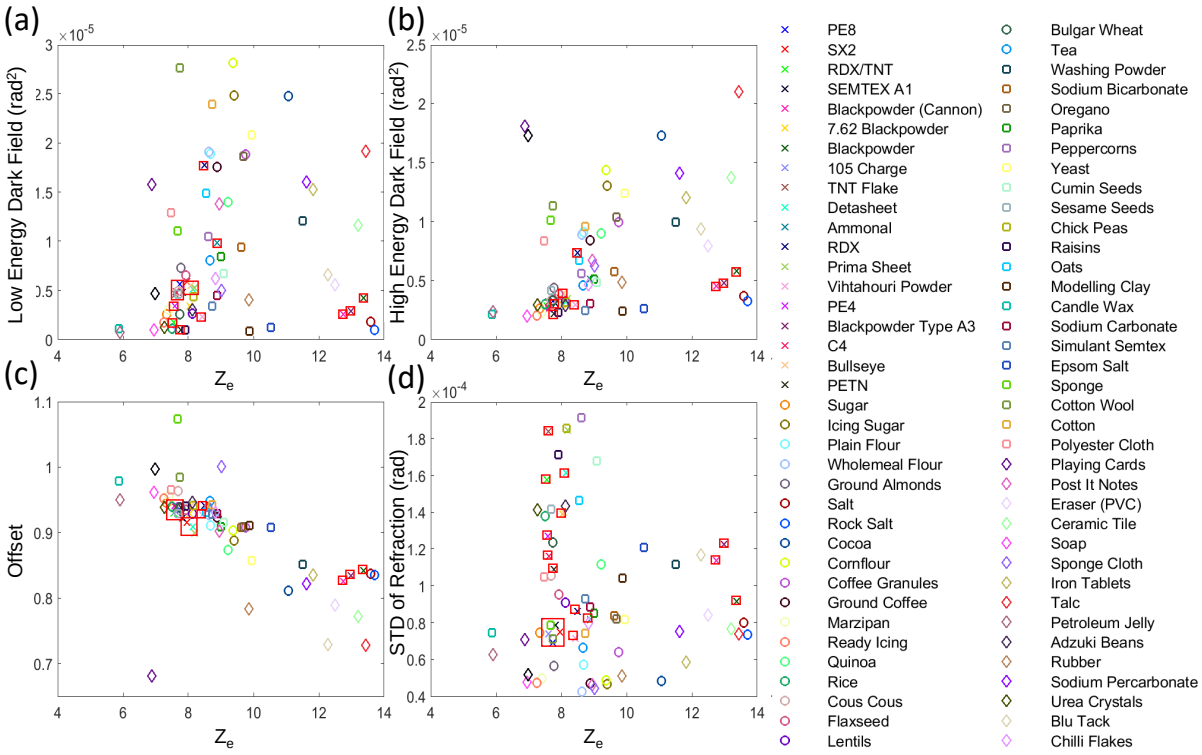

Fig. 2. Scatterplots combining the effective atomic number (Z_e, horizontal axis in all graphs) with dark-field at low and high energy, offset, and standard deviation of refraction, reported on the vertical axis of panels (a)–(d), respectively. Explosive materials are highlighted by squares.

of an unknown sample being made of a certain material as the overlap between hyper-cubes (hyper-spheres).

Those probabilities can be used to produce user-friendly, "material-specific" images, an example of which is provided in Fig. 4; the sides (radii) of the hyper-cubes (spheres) can be used to trade off sensitivity versus specificity. The examples in Figs. 4(a)–4(d) and 4(e)–4(h) refer to bicarbonate of soda and semtex, respectively. Figures 4(a) and 4(b) and Figs. 4(e), 4(f) use Z_e only, with a smaller [Figs. 4(a) and 4(e)] and larger [Figs. 4(b) and 4(f) thresholds resulting in high specificity and sensitivity, respectively.

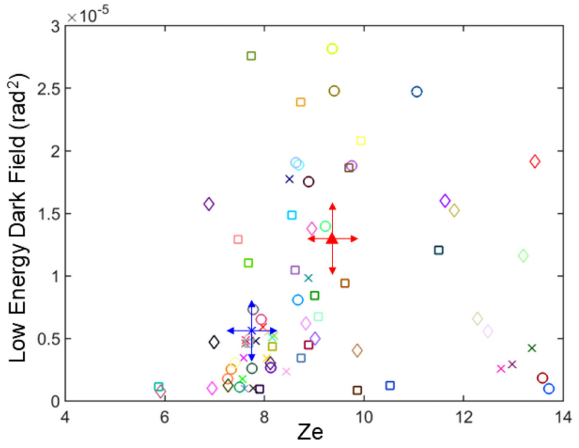

Fig. 3. Material discrimination based on relative distances in "image contrast" space, with the overlap between areas defined by the uncertainty on the measured contrasts determining the probability of an unknown (red) material corresponding a known one (blue). While only two contrasts can be represented in a 2D figure, in truth the overlap between "hyper-volumes" is calculated by simultaneously considering the uncertainties on all contrast channels.

Figures 4(c) and 4(d) and Figs. 4(g) and 4(h) introduce dark-field at two energies, offset and standard deviation of refraction. Multiple contrast channels lead to multiple ways to trade-off sensitivity and specificity: for example, highly specific images [Figs. 4(c) and 4(g)] require that Z_e is triggered alongside three of the other four contrast channels, while Figs. 4(d) and 4(h) simply require that three channels are flagged, one of which may or may not be Z_e. As can be seen the multi-contrast approach works extremely well for the bicarbonate of soda, where it completely clears up the image; relaxing the specificity makes almost all pixels in the bicarbonate ROI turn yellow, at the expense of some spurious positive pixels in other parts of the image. Things work slightly less well for semtex; however, the combination of contrasts significantly outperforms the corresponding Z_e only cases. In future work, outcomes could be further improved by (a) giving different weights to the various contrast channels, which could be refined with successive and extensive calibrations and (b) performing additional probability calculations based on the number of pixels flagged in a ROI, e.g., by assuming "continuity" for the examined material and defining likely boundaries. Supplement 1, Fig. 1 provides additional examples for different materials, and Supplement 1, Figs. 2 and 3 show that the method is still applicable when different materials overlap. Application of the above principle to a range of different materials allows the creation of confusion matrices, an example of which is provided in Fig. 5.

Confusion matrices like those of Fig. 5 enable presenting results like those of Fig. 4 for a plurality of materials in a single image. The gray scale represents the probability that a material belongs to a certain class, with white and black representing the certainty of belonging/not belonging to that class, respectively. The same materials are listed along the horizontal and vertical axis so that a

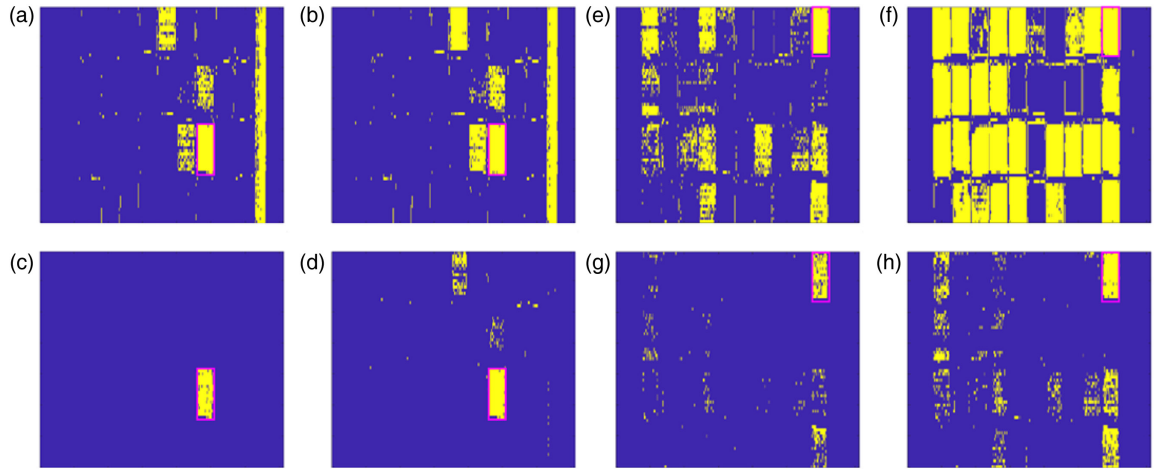

Fig. 4. Material-specific images for (a)–(d) the bicarbonate of soda and (e)–(h) semtex, the area corresponding to which is highlighted by a pink rectangle in all images. All images are binary, flagging the pixels as belonging to the material in question (yellow) or not (blue). Images (a), (b) and (e), (f) are based only on Z_e, with (a), (e) short and (b), (f) long cutoff radii corresponding to high specificity and high sensitivity conditions, respectively. Images (c), (d) and (g), (h) include the additional contrast channels introduced in Fig. 2. The plurality of contrasts provides different mechanisms to trade-off specificity and sensitivity; for example, in images (c) and (g) (high specificity), Z_e had to be triggered alongside 3 of the other 4 contrast channels for a pixel to be flagged as belonging to a certain material. Conversely, "high sensitivity" images in (d) and (h) simply require that 3 contrasts out of 5 are triggered, regardless of Z_e. The improvement brought by the proposed multi-contrast method can be appreciated by comparing each image of the bottom row to its "Z_e only" correspondent immediately above.

white diagonal in a black background represents perfect detection with 100% certainty (here we have represented twice as many materials on the vertical than on the horizontal axis, so the diagonal would only occupy half the graph). As can be seen, the Z_e cases in Figs. 5(a) and 5(b) are severely plagued by false positives, also in the high specificity case. The multi-contrast analysis significantly clears up the confusion matrix; it must be noted, however, that the diagonal points in Fig. 5(c) are dimmer than in Figs. 5(a) and 5(b). This is ameliorated by relaxing the specificity in favor of sensitivity [Fig. 5(d)], which results in the appearance of a larger number

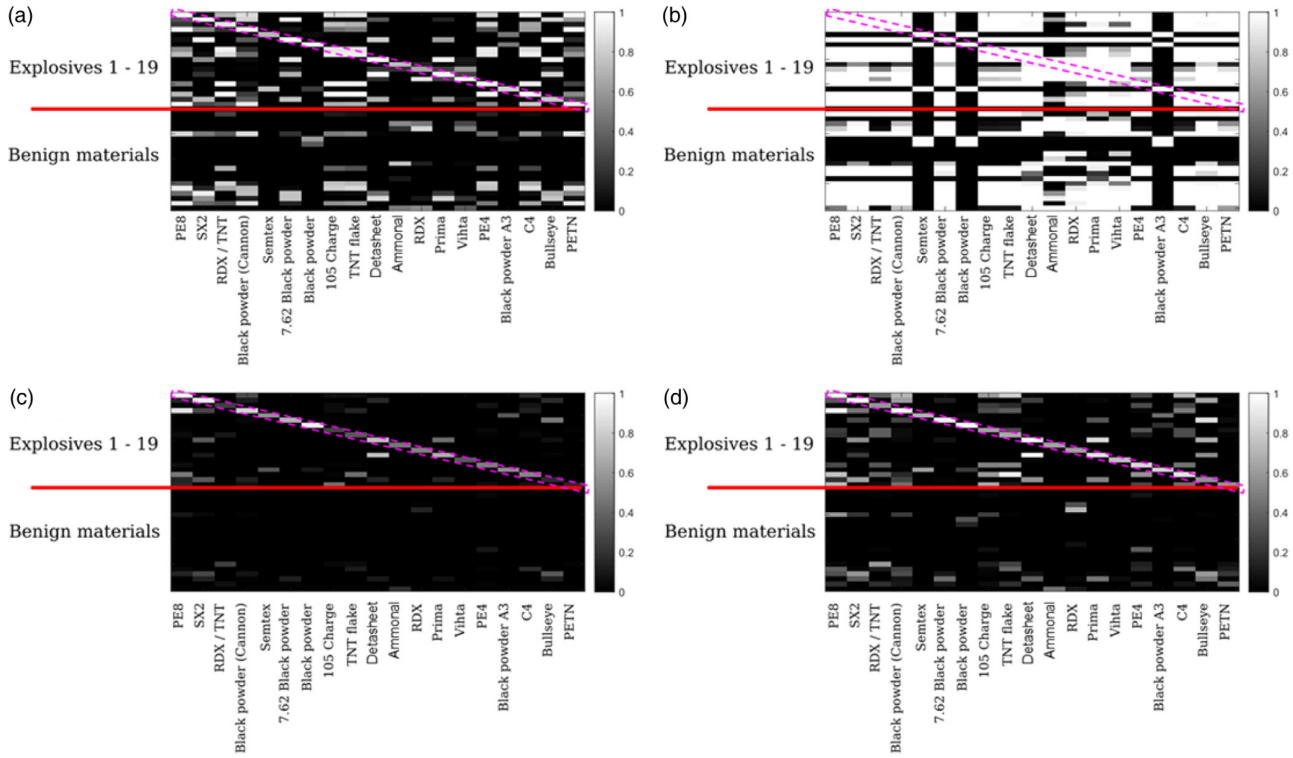

Fig. 5. Confusion matrices based on Z_e at (a) high specificity and (b) high sensitivity, and on the multi-contrast approach at (c) high specificity and (d) high sensitivity. Matrices focus only on the explosive materials, which are listed on the horizontal axis. The same materials in the same order occupy the top 19 rows of each matrix, while the bottom 19 represent a subset of the benign materials. The gray scale represents the probability that a certain material belongs to a certain class; hence, the ideal matrix is a white diagonal occupying the top half of the graph, with all other entries being black.

Table 1. Results of Machine Learning Analysis Performed with Various Architectures[a]

Architecture	TP	FP	FN	Precision	Recall
Single softmax	297	65	16	82.04%	94.89%
Hierarch. softmax	303	57	10	84.17%	96.81%
Hierarch. x-entropy	307	52	6	85.52%	98.08%
Hierarch. x-entropy + layer-specific learn. adapt.	312	32	1	90.70%	99.68%
Hierarch. x-entropy + layer-specific learn. adapt. **no dark-field**	281	56	32	83.38%	89.78%

[a]The first four rows refer to datasets combining attenuation at two energies and dark-field at two energies, while the two dark-field images were removed to obtain the results of row 5. TP = true positives, FP = false positives, FN = false negatives, Precision = TP/(TP + FP), Recall = TP/(TP + FN).

of false positives. Overall, Fig. 5(d) has a comparable diagonal to Fig. 5(a), but fewer (and dimmer) false positives.

Finally, we present the results of a test carried out using deep learning. For simplicity, we focused on the performance of attenuation at high and low energy alone vs. their combination with dark-field at high and low energy. A random combination of materials with random thickness ranging from 12 to 30 mm was placed inside bags with additional cluttering items; ratio images were used to eliminate the dependence from material thickness as described in [33]. A total of 3891 combinations were produced and scanned with the imaging system described in Section 2. Supplement 1, Table 1 lists all threat and non-threat materials as well as all cluttering items, Supplement 1, Fig. 4 provides a scheme summarizing the characteristics of the dataset used in the deep learning experiments, Supplement 1, Fig. 5 shows a few example images from the scanning campaign.

The 3891 images were split into 2732 (70%) for training and 1168 (30%) for testing. The convolutional neural network (CNN) Inception V3 [47] was used as our base architecture, pre-trained on ImageNet [48]; Supplement 1, Fig. 6 shows a schematic of our hierarchical architecture (that yielded the best performance), and more details on it are provided in the methods section. We progressively refined our approach by (1) moving from a single to a hierarchical architecture in which the cluttering object classes are segregated first then the materials differentiated, (2) swapping the softmax for a cross-entropy loss function, and (3) implementing layer-specific learning rate adaptation [45]: every step increased the obtained precision and recall values. To calculate these, we focused on the 313 images out of the 1168 used for testing that contained an explosive. We counted the number of true positives (TP), false positives (FP), and false negatives (FN), and calculated precision as TP/(TP + FP) and recall as TP/(TP + FN); note that TP + FN = 313 while typically TP + FP > 313 as FPs are images outside the 313 explosive-containing ones erroneously assumed to contain an explosive. Precision and recall provide estimates of specificity and sensitivity, respectively; results are reported in Table 1 for the various architecture refinements. The best performing architecture (hierarchical segregation + cross-entropy loss function + layer-specific learning rate adaptation) was also applied to a dataset from which the dark-field images had been eliminated, resulting in a significantly reduced performance.

The table clearly shows the improvement in both precision and recall values as the deep learning architecture is progressively refined, reaching a near-perfect 99.68% degree of recall for the hierarchical architecture with cross-entropy loss function and layer-specific learning adaptation; note this corresponds to a single false negative out of the 313 explosive-bearing images included in the 1168 used for testing. Importantly, removing the dark-field imaging channel from the dataset leads to a significant degradation in the overall performance, with a reduction of 7% and 10% points in the precision and recall values, respectively, thereby demonstrating the importance of the dark-field channel in the material discrimination potential of deep neural networks.

4. CONCLUSION

We have shown that the inclusion of additional x-ray contrast channels, namely dark-field at two average energies, attenuation at a third energy and standard deviation of refraction, can significantly enhance the material discrimination and identification potential of dual-energy x-ray imaging. We developed a mechanism to combine all the different contrast channels into a single "material-specific" image, the pixels of which represent whether they belong to that material based on overlapping hyper-spheres, the radii of which are a chosen uncertainty. We propose a mechanism to trade-off sensitivity and specificity in this approach and compare the results to those obtained while using only attenuation at two x-ray energies, demonstrating substantial advantages. We also provide evidence that the method can be extrapolated to overlapping materials, so long as areas where we can "re-normalize" the signal against the overlapping object/background are available.

This laid the groundwork for the application of deep learning to the multi-contrast images, which was previously shown to bear promise in a proof-of-concept study involving a small dataset and a limited number of materials [33]. In this wider study, a custom-developed architecture was applied to almost 4000 images containing 57 benign and 19 explosive materials with varying thickness, randomly mixed in bags also containing a multitude of cluttering objects. The network architecture was progressively refined by creating a hierarchical structure in which the cluttering object classes are segregated first then materials are discriminated, as well as by introducing a cross-entropy loss function and layer-specific learning rate adaptation; this ultimately led to a near-perfect recall rate of 99.68%, corresponding to a single explosive being missed out of the 313 contained in the 1168 images used for testing. By applying the same deep neural network to the same image set from which the dark-field images were excluded, the recall rate dropped by 10 percentage points, corresponding to 32 explosives being missed. We find this highlights the fundamental importance of the match between dark-field images and detection abilities of the deep neural network in discriminating certain material classes. Although security inspections have been chosen as an example to demonstrate the potential of the approach, the method is purely based on the additional discrimination capabilities provided by the textural nature of dark-field images, related to the microscopic structure of a given material; as such, it can be widely applied to microscopically inhomogeneous materials across the life and physical sciences.

In terms of limitations, it has to be mentioned that, despite the increase compared to previous studies like the one described in [33], ∼4000 is still a small number of images for a typical deep learning study. This is largely due to the fact that we are concurrently proposing a new x-ray imaging method currently at the pre-commercial prototype stage, which means there are no larger image databases available besides those generated by the method's inventors and/or developers. To provide at least some reassurance on the dataset's diversity, we have analyzed the dataset's variability by studying the standard deviations of the 2048-dimensional feature vectors extracted from the fully connected penultimate (pre-classification) layer of the Inception V3 network (see Supplement 1, Fig. 7 and related discussion). To test its representativeness, we have compared the range of textures present in our dataset to those of a widely used "real world" one, the Oxford's Describable Textures Dataset [53], with results presented in Supplement 1, Fig. 8. We have also tested the method's robustness by applying our best-performing CNN architecture to images degraded through convolution with Gaussians with increasingly large standard deviations. The results (presented in Supplement 1, Table 2) show that, although small standard deviations have a limited effect on performance, this degrades more significantly as the standard deviation is made larger. While this may indicate some limits in the approach's robustness, it also provides indications for future work by, e.g., fine-tuning our model as a downstream task over a self-supervised trained pretext model [54] and preferably using transformers throughout [55,56].

Funding. Royal Academy of Engineering (CiET1819/2/78); Home Office (84887-537251); Engineering and Physical Sciences Research Council (EP/T005408/1).

Acknowledgment. A. O. was supported by the Royal Academy of Engineering under the Chairs in Emerging Technologies scheme.

Disclosures. D. B. is a Nikon employee; A. A. was a Nikon employee at the time this research was conducted. A. O. is a named inventor on patents held by UCL protecting the x-ray imaging technology used to obtain the results presented in this paper.

Data availability. The datasets generated and/or analyzed during the current study are not publicly available on grounds of security but are available from the corresponding author on request. The computer code used to perform phase retrieval is available from A. A. The machine learning code is available from S. S. S. Interested readers are invited to contact the corresponding author.

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. U. Bonse and M. Hart, "An x-ray interferometer," Appl. Phys. Lett. **6**, 155–156 (1965).
2. E. Förster, K. Goetz, and P. Zaumseil, "Double crystal diffractometry for the characterization of targets for laser fusion experiments," Kristall Tech. **15**, 937–945 (1980).
3. A. Snigirev, I. Snigireva, V. Kohn, *et al.*, "On the possibilities of x-ray phase-contrast microimaging by coherent high-energy synchrotron radiation," Rev. Sci. Instrum. **66**, 5486–5492 (1995).
4. A. Momose, T. Takeda, Y. Itai, *et al.*, "Phase-contrast X-ray computed tomography for observing biological soft tissues," Nat. Med. **2**, 473–475 (1996).
5. T. J. Davis, D. Gao, T. E. Gureyev, *et al.*, "Phase-contrast imaging of weakly absorbing materials using hard x-rays," Nature **373**, 595–598 (1995).
6. S. W. Wilkins, T. E. Gureyev, S. Gao, *et al.*, "Phase-contrast imaging using polychromatic hard x-rays," Nature **384**, 335–338 (1996).
7. D. Chapman, W. Thomlinson, R. E. Johnston, *et al.*, "Diffraction enhanced x-ray imaging," Phys. Med. Biol. **42**, 2015–2025 (1997).
8. P. Cloetens, W. Ludwig, J. Baruchel, *et al.*, "Holotomography: quantitative phase tomography with micrometer resolution using hard synchrotron radiation x-rays," Appl. Phys. Lett. **75**, 2912–2914 (1999).
9. A. Olivo, F. Arfelli, G. Cantatore, *et al.*, "An innovative digital imaging set-up allowing a low-dose approach to phase contrast applications in the medical field," Med. Phys. **28**, 1610–1619 (2001).
10. C. David, B. Nohammer, and H. H. Solak, "Differential x-ray phase contrast imaging using a shearing interferometer," Appl. Phys. Lett. **81**, 3287–3289 (2002).
11. K. S. Morgan, D. M. Paganin, and K. K. W. Siu, "Quantitative single-exposure x-ray phase contrast imaging using a single attenuation grid," Opt. Express **19**, 19781–19789 (2011).
12. K. S. Morgan, D. M. Paganin, and K. K. W. Siu, "X-ray phase imaging with a paper analyser," Appl. Phys. Lett. **100**, 124102 (2012).
13. S. Berujon, H. Wang, and K. Sawhney, "X-ray multimodal imaging using a random-phase object," Phys. Rev. A **86**, 063813 (2012).
14. L. Rigon, H. J. Besch, F. Arfelli, *et al.*, "A new DEI algorithm capable of investigating sub-pixel structures," J. Phys. D Appl. Phys. **36**, A107 (2003).
15. O. Oltulu, Z. Zhong, M. Hasnah, *et al.*, "Extraction of extinction, refraction and absorption properties in diffraction enhanced imaging," J. Phys. D Appl. Phys. **36**, 2152–2156 (2003).
16. E. Pagot, P. Cloetens, S. Fiedler, *et al.*, "A method to extract quantitative information in analyzer-based x-ray phase contrast imaging," Appl. Phys. Lett. **82**, 3421–3423 (2003).
17. F. Pfeiffer, M. Bech, O. Bunk, *et al.*, "Hard-x-ray dark-field imaging using a grating interferometer," Nat. Mater. **7**, 134–137 (2008).
18. M. Endrizzi, P. C. Diemoz, T. P. Millard, *et al.*, "Hard x-ray dark-field imaging with incoherent sample illumination," Appl. Phys. Lett. **104**, 024106 (2014).
19. I. Zanette, T. Zhou, A. Burvall, *et al.*, "Speckle-based x-ray phase-contrast and dark-field imaging with a laboratory source," Phys. Rev. Lett. **112**, 253903 (2014).
20. F. Pfeiffer, T. Weitkamp, O. Bunk, *et al.*, "Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources," Nat. Phys. **2**, 258–261 (2006).
21. A. Olivo and R. Speller, "A coded-aperture technique allowing x-ray phase contrast imaging with conventional sources," Appl. Phys. Lett. **91**, 074106 (2007).
22. T. E. Gureyev, D. M. Paganin, B. Arhatari, *et al.*, "Dark-field signal extraction in propagation-based phase-contrast imaging," Phys. Med. Biol. **65**, 215029 (2020).
23. K. Willer, A. A. Fingerle, W. Noichl, *et al.*, "X-ray dark-field chest imaging for detection and quantification of emphysema in patients with chronic obstructive pulmonary disease: a diagnostic accuracy study," Lancet Digital Health **3**, e733–e744 (2021).
24. P. C. Diemoz, F. A. Vittoria, and A. Olivo, "Spatial resolution of edge illumination X-ray phase-contrast imaging," Opt. Express **22**, 15514–15529 (2014).
25. I. Buchanan, S. Cipiccia, C. Peiffer, *et al.*, "Direct x-ray scattering measurements in edge-illumination/beam-tracking and their interplay with the variance of the refraction signals," Phys. Rev. Appl. **10**, 041406 (2023).
26. D. Pacella, "Energy-resolved x-ray detectors: the future of diagnostic imaging," Rep. Med. Imaging **8**, 1–13 (2015).
27. E. Roessl and R. Proksa, "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," Phys. Med. Biol. **52**, 4679–4696 (2007).
28. V. Sriskaran, J. Alozy, T. Ballabriga, *et al.*, "High-rate, high-resolution single photon X-ray imaging: Medipix 4, a large 4-side buttable pixel readout chip with high granularity and spectroscopic capabilities," J. Instrum. **19**, P02024 (2024).
29. L. Jowitt, M. Wilson, P. Seller, *et al.*, "HEXITEC 2 × 2 tiled hard X-ray spectroscopic imaging detector system," J. Instrum. **17**, P01012 (2022).
30. O. Scharf, S. Ihle, I. Ordavo, *et al.*, "Compact pnCCD-based X-ray camera with high spatial and energy resolution: a color X-ray camera," Anal. Chem. **83**, 2532–2538 (2011).
31. M. N. Boone, F. Van Assche, S. Vanheule, *et al.*, "Full-field spectroscopic measurement of the X-ray beam from a multilayer monochromator using a hyperspectral X-ray camera," J. Synchrotron Rad. **27**, 110–118 (2020).
32. A. Castoldi, E. Gatti, C. Guazzoni, *et al.*, "The controlled-drift detector," Nucl. Instrum. Meth. Phys. Res. A **439**, 519–528 (2000).

33. T. Partridge, A. Astolfo, S. S. Shankar, *et al.*, "Enhanced detection of threat materials by dark-field x-ray imaging combined with deep neural networks," Nat. Commun. **13**, 4651 (2022).
34. R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerized tomography," Phys. Med. Biol. **21**, 733–744 (1976).
35. T. Sellerer, K. Mechlem, R. Tang, *et al.*, "Dual-energy x-ray dark-field material decomposition," IEEE Trans. Med. Imaging **40**, 974–985 (2021).
36. X. Wu, H. Liu, and A. Yan, "X-ray phase-attenuation duality and phase retrieval," Opt. Lett. **30**, 379–381 (2005).
37. S. V. Naydenov and V. D. Ryzhikov, "Multi-energy techniques for radiographic monitoring of chemical composition," Nucl. Instrum. Meth. Phys. Res. A **505**, 556–558 (2003).
38. A. Astolfo, I. G. Haig, D. Bate, *et al.*, "Increased material differentiation through multi-contrast x-ray imaging: a preliminary evaluation of potential applications to the detection of threat materials," Phys. Scr. **98**, 095501 (2023).
39. K. Wells and D. Bradley, "A review of X-ray explosives detection techniques for checked baggage," Appl. Radiat. Isot. **70**, 1729–1746 (2012).
40. S. U. Khan, I. U. Khan, I. Ullah, *et al.*, "A review of airport dual energy x-ray baggage inspection techniques: image enhancement and noise reduction," J. X-Ray Sci. Technol. **28**, 481–505 (2020).
41. A. Shevchuk, J. P. O. Evans, A. J. Dicken, *et al.*, "Combined x-ray diffraction and absorption tomography using a conical shell beam," Opt. Express **27**, 21092–21101 (2019).
42. A. Alzubaidi, J. Zhang, A. J. Humaidi, *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," J. Big Data **8**, 53 (2021).
43. S. Akcay and T. P. Breckon, "Towards automatic threat detection: a survey of advances of deep learning within X-ray security imaging," Pattern Recognit. **122**, 108245 (2022).
44. A. Mouton and T. P. Breckon, "A review of automated image understanding within 3D baggage computed tomography security screening," J. X-Ray Sci. Technol. **23**, 531–555 (2015).
45. S. Shankar, D. Robertson, Y. Ioannou, *et al.*, "Refining architectures of deep convolutional neural networks," in *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition* (2016), Vol. **7780612**, pp. 2212–2220.
46. A. Astolfo, I. Buchanan, T. Partridge, *et al.*, "The effect of a variable focal spot size on the contrast channels retrieved in edge-illumination x-ray phase contrast imaging," Sci. Rep. **12**, 3354 (2022).
47. C. Szegedy, V. Vanhoucke, S. Ioffe, *et al.*, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 2818–2826.
48. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (2012), pp. 1097–1105.
49. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning* (2015), pp. 448–456.
50. A. F. Agarap, "Deep learning using rectified linear units (ReLU)," arXiv, arXiv:1803.08375 (2018).
51. S. G. Azevedo, H. E. Martz, Jr., M. B. Aufderheide, III, *et al.*, "System-independent characterization of materials using dual-energy computed tomography," IEEE Trans. Nucl. Sci. **63**, 341–349 (2016).
52. D. Shoukroun, L. Massimi, M. Endrizzi, *et al.*, "Composite porosity characterization using x-ray edge illumination phase contrast and ultrasonic techniques," Proc. SPIE **11593**, 115932M (2021).
53. M. Cimpoi, S. Maji, I. Kokkinos, *et al.*, "Describing textures in the wild," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2014), pp. 3606–3613.
54. T. Chen, S. Kornblith, M. Norouzi, *et al.*, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning* (PMRL, 2020), pp. 1597–1607.
55. K. He, X. Chen, S. Xie, *et al.*, "Masked autoencoders are scalable vision learners," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2022), pp. 16000–16009.
56. D. Zhang and F. Zhou, "Self-supervised image denoising for real-world images with context-aware transformer," IEEE Access **11**, 14340–14349 (2023).